

ABCI 3.0開発加速利用（2024年度）成果概要（公開用）

課題名：ABCI3.0上でのマルチモーダル事前学習のための共通基盤開発	実施時期： 所属機関名：東京科学大学 代表者氏名：横田理央
成果概要：ABCI3.0上でMegatron-LMをベースとしたマルチモーダルモデルに対応可能なフレームワークを整備した。Molmo, Idefics3, MVLM-Dなどの画像・言語モデル(VLM)やLlama3.1+Whisperからなる音声・言語モデル(AudioLM)をMegatron-LMから利用できるようにした。2024年11月の試験利用ではH200あたり100TFLOP/s程度の性能にとどまっていたのを400TFLOP/sまで向上した。	
成果のポイント：VLMにおける継続事前学習の可能性について検討するために、Molmo-1B, 7B, 72B Idefics3-8B NVLM-D-72Bのモデルについて日本語Visual Instruction Tuningおよび日本語Visual継続事前学習を行った。Megatron-LMおよび独自ライブラリvlm-recipesを利用して学習を行うが、LLMとは異なりVLMの効率的な学習には実装面で課題が存在する。そのため、LLMにおいては440～500 TFLOP/s で学習できる設定を用いても、VLMでは1/5～1/3程度の学習速度でしか学習することができない。また、本実験に備え、Learning Rateの探索、学習データセットのアブレーションのために各モデルごとに10程度の予備実験を行った。学習を行うためのライブラリがLLaVAやframingoアーキテクチャのみに対応しており、ライブラリ側の改修が必要であり、そのためのデバッグ及び改修を行った後の検証を行う必要があった。AudioLM における言語モデルを活用した事前学習の可能性について検討するために、Swallow を事前学習モデルとして日本語 Audio Pre-training 及び Instruction Tuning を行う。独自ライブラリ audiolm-recipes を利用して学習を行うが、LLM とは異なり AudioLM の効率的な学習には実装面で課題が存在する。そのため、LLMにおいては 440～500 TFLOP/s で学習できる設定を用いても、(VLM と同様に) AudioLM では1/5～1/3程度の学習速度でしか学習することができない。また、本実験に備え、Learning Rate の探索、学習データセットのアブレーションのために 10 程度の予備実験を行った。さらに、AudioLM の効率的な学習を実現するため、網羅的な分散並列学習設定の探索と、AudioLM の学習に対応した学習支援ツールを作成した。	
成果についてより詳細な情報を提供しているWebページ、発表論文などの情報： Webページはなし 論文は準備中	